Modeling the Chemistry of Complex Petroleum Mixtures

Richard J. Quann

Mobil Technology Company, Paulsboro, New Jersey

Determining the complete molecular composition of petroleum and its refined products is not feasible with current analytical techniques because of the astronomical number of molecular components. Modeling the composition and behavior of such complex mixtures in refinery processes has accordingly evolved along a simplifying concept called lumping. Lumping reduces the complexity of the problem to a manageable form by grouping the entire set of molecular components into a handful of lumps. This traditional approach does not have a molecular basis and therefore excludes important aspects of process chemistry and molecular property fundamentals from the model's formulation. A new approach called structure-oriented lumping has been developed to model the composition and chemistry of complex mixtures at a molecular level. The central concept is to represent an individual molecule or a set of closely related isomers as a mathematical construct of certain specific and repeating structural groups. A complex mixture such as petroleum can then be represented as thousands of distinct molecular components, each having a mathematical identity. This enables the automated construction of large complex reaction networks with tens of thousands of specific reactions for simulating the chemistry of complex mixtures. Further, the method provides a convenient framework for incorporating molecular physical property correlations, existing group contribution methods, molecular thermodynamic properties, and the structure–activity relationships of chemical kinetics in the development of models. — *Environ Health Perspect* 106(Suppl 6):1441–1448 (1998). *http://ehpnet1.niehs.nih.gov/docs/1998/Suppl-6/1441-1448quann/abstract.html*

Key words: complex mixtures, structure-oriented lumping, petroleum, mathematical modeling, structure–activity relationships, chemical kinetics, molecular thermodynamics

Many chemical reaction systems found in industry and nature are complex in terms of the large number of molecular components and chemical reactions. Examples include petroleum refining, biologic systems, and combustion processes. The number of components can exceed 10^4 and the number of chemical reactions can be an order of magnitude greater. Developing models of such large reacting systems is difficult because of the scale and the lack of fundamental information. Analytical techniques cannot identify and measure all the individual molecular components when so many are present a mixture. It is also impractical to study the mechanisms and kinetics of all possible chemical reactions.

Further, how could this information be managed, organized, and formulated into a model simulating the reactions of so many components? Accordingly, models developed for complex reacting systems have used a technique called lumping (1,2) to simplify the representation of composition and chemistry.

Lumping partitions the entire molecular population into a small number of lumps (approximately 10) determined by the limitations in measuring mixture composition, the similar chemistry and physical properties of molecular components within a lump, and the model's intended purpose. This approach reduces complexity and eliminates the need for a detailed molecular characterization, albeit at the expense of its usefulness. Using petroleum refining as an example, the model shown in Figure 1 was developed in the 1970s to predict gasoline yield for the fluid catalytic cracking process that converts heavy gas oil containing high-molecular-weight hydrocarbons to the lower molecular weight components of gasoline. It represents the complex composition of gas oil as eight lumps based on the crude analytical capabilities of that time—broad

molecular categories and boiling range. Although this model can predict gasoline yield from these lumps, it is not structured to predict gasoline composition or the impact of composition on the required quality specifications. In addition, the gasoline yield is not reliably predicted because the lumping scheme cannot represent other important variations in gas oil composition and because the process chemistry is inadequately represented.

Models developed for refining processes must have the capability of predicting not just accurate product yields but also the product's molecular composition, the boiling point distribution, certain mixture physical properties and quality specifications that depend on composition, and the plant's operating conditions. A molecular approach, if possible, could satisfy these requirements by incorporating more detailed petroleum characterization, process chemistry, and molecular property correlations. Recently developed analytical methods are capable of probing some aspects of petroleum's molecular structure and composition, and there is an extensive, but not complete, knowledge base of the chemistry of refining processes from laboratory studies of selected hydrocarbons. A recently developed approach called structure-oriented lumping (SOL) attempts to model the complex composition and chemistry of petroleum mixtures at the molecular level (3). This paper reviews the concepts of SOL and the strategy for developing molecular-based models of reaction systems where the number of molecular components is very large and their molecular structure cannot be completely determined.

Organizing the Composition of Petroleum

The basic concept of the SOL approach is that any hydrocarbon molecule can be constructed from a set of different structural features or increments. A structural increment is a specific combination or configuration of C, H, S, N, and O atoms, often occurring in different molecules. The set of 22 increments is shown in Figure 2. The increments consist of three types of aromatic rings (A6, A4, A2), six types of naphthenic rings (N6–N1), a methylene $-CH_2-$ group (R), bridging between rings (A–A), hydrogen deficiency (H), heteroatom structures containing S, N, and O, and the degree of branching (br) and ring

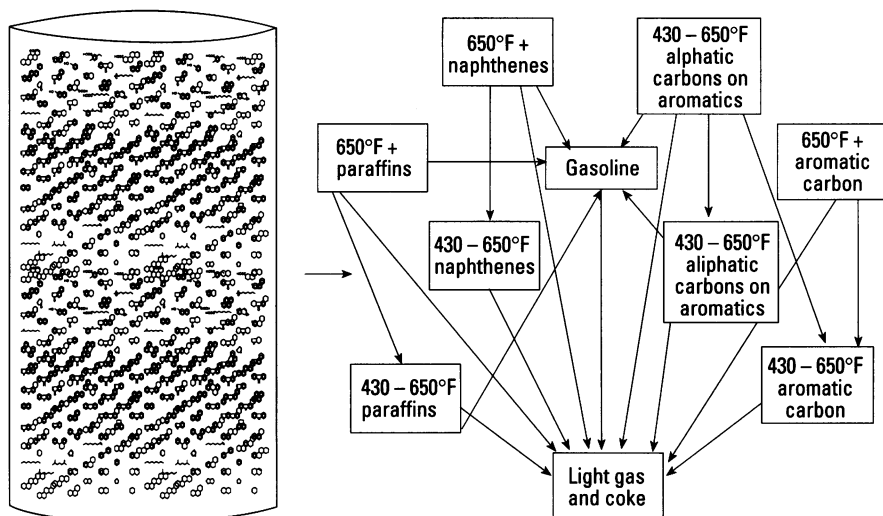Figure 1. Modeling the fluid catalytic cracking process using a simplified representation of petroleum's complex composition.
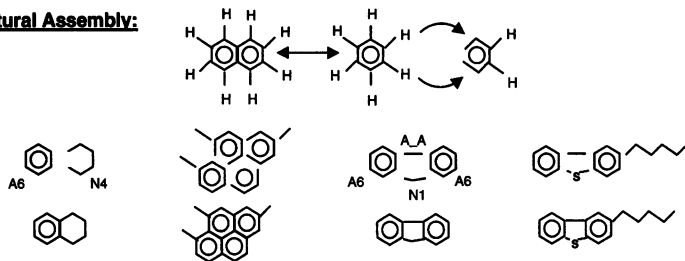

Figure 2. Structural increments and their stoichiometry for constructing molecules. Modified from Quann and Jaffe (3).

substitutions (me). Each increment has a specific C, H, S, N, and O stoichiometry and thus molecular weight. The term increment is applicable because most cannot exist independently but must occur as an incremental part of a molecule.

The SOL method mathematically organizes this set of increments into a vector, with each element of the vector corresponding to one of the increments. A molecule is represented as a 22-element numerical vector, with each molecule distinguished by different numbers and types of increments, as shown in Figure 3.

For example, a vector with a one in the first element or position (A6) and zero for all others represents a benzene molecule. Naphthalene is constructed with A6 and A4 increments, having one for the first two elements of the vector and zero for all others. Other examples are shown in Figure 3. Molecular elemental stoichiometry and molecular weight are easily computed from the contributions of the increment's elemental stoichiometry.

Any molecule found in petroleum can be represented by this method, altough not necessarily uniquely. Many molecules may

have identical vector representation. The increments count structures present in the molecule. Increment spatial orientation is not represented by the structure vector. Molecules having the same structure vector but whose increments are arranged differently are called structural isomers. Examples of structural isomer sets that have identical vector representations but with clearly different spatial arrangement of the increments are shown in Figure 4. The br and me increments do not contribute to the molecule's stoichiometry but are used for a limited description of certain isomeric features: the number of branches on a paraffin or alkyl chain and the number of methyl substituents on rings, respectively. Again, the method as shown here does not assign specific locations for these branches or ring substitutions, only the total number for the molecule. For example, there are only 21 possible structure vector representations for all possible C_{43} paraffins, yet there are over a thousand-billion isomeric permutations. Not distinguishing structural isomers is a practical consequence of analytical limitations, but it also has other implications, as we will discuss.

A complex mixture in SOL is represented as a set of vectors with each vector corresponding to a molecule or an ensemble of structural isomers. The composition of the mixture is expressed as the weight or mole percent of each vector. It is helpful to visualize the mixture using the concepts of molecular class and homologous series. Some common molecular classes found in petroleum or refinery products are shown (without alkyl substituents) in Figure 5. They are all described by the SOL method and have different combinations of structural increments. A homologous series is composed of molecules from the same molecular class, varying only in R, br, and me because of the presence of alkyl substituents. The number for R is the total number of carbons for a paraffin or olefin or on the alkyl substituents of a ring compound. The simplest example of a homologous series in petroleum is the normal paraffins (br = me = 0) extending from methane (R = 1) to over 60 carbon atoms. Other homologous series are more complicated in that br and me may vary with R to represent some average branching and substitution for the set of isomers at each carbon number of each molecular class. A complex petroleum mixture in SOL is thus composed of some 5000 to 10,000 vectors—one for each molecule or isomer ensemble—and organized into

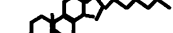|	A6	A4	A2	N6	N5	N4	N3	N2	N1	R	br	me	H	A_A	S	RS	AN	NN	RN	O	RO	O=
2,3,4-Trimethylpentane	0	0	0	0	0	0	0	0	0	8	3	0	1	0	0	0	0	0	0	0	0	0
Benzene	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Naphthalene	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Phenanthrene	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pyrene	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Carbazol	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0
Benzofuran	1	0	0	0	0	1	0	0	0	0	0	0	−1	0	0	0	0	0	0	1	0	0
Fluoranone	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1
Adamantane	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Cholesterol	0	0	0	1	0	2	1	0	0	10	2	2	−1	0	0	0	0	0	0	0	1	0
Water H_2O	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0

Figure 3. Examples of the structure-oriented lumping vector representation of molecular structure. Modified from Quann and Jaffe (3).
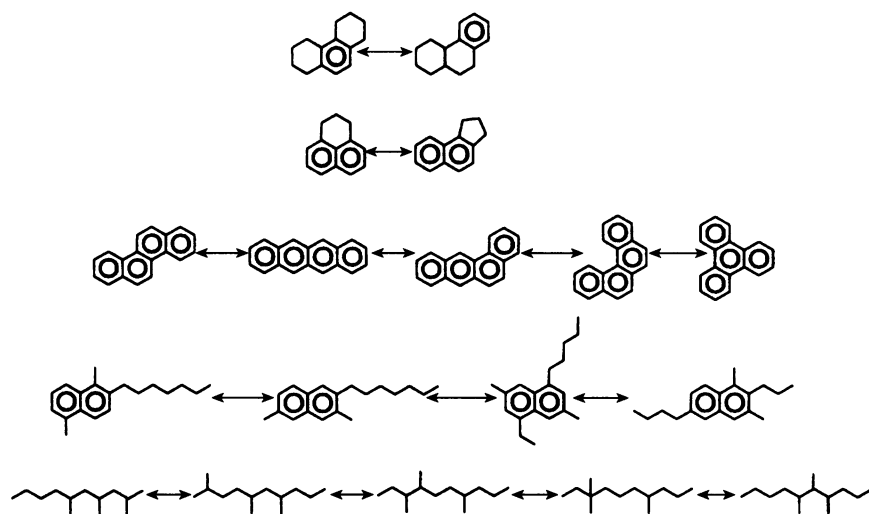
Figure 4. Examples of isomers having the same structure-oriented lumping vector representation.

about 150 molecular classes and their homologous series.

Describing molecular structure at this level is consistent with our ability to analyze petroleum. Modern gas chromatography techniques routinely identify and quantify most of the several hundred components, including isomers, of the low-boiling (< 450°K), gasoline-range material in petroleum or products. However, the numerous possible increment spatial orientations result in far too many components (isomers) for analytical techniques to identify in specific molecular structures that boil above 450°K. The only measurement is the number of molecules contributing to a molecular mass. By combining chromatographic separations with mass spectrometry (4), the most likely molecular class and the total carbon number (R) of alkyl groups are inferred at each molecular mass, providing the homologous series distribution of each molecular class.

The structure of the alkyl groups and all their possible permutations cannot be determined at each mass, with the possible exception of n-alkyls. An average or archetype structural arrangement of alkyl groups can be inferred from nuclear magnetic resonance measurements of the mixture but this does not necessarily imply a common arrangement. Similarly, the spatial arrangement of ring structures can have many permutations. The biologic origin of petroleum would suggest that the cata configuration found in terpenes and steroids is the likely isomeric form of the multiring structure (5). Hence, the SOL description of molecular structure is consistent with the limitations of current analytical methods. The SOL approach assigns a molecular structure to the large number (approximately 5000) of distinguishable

Figure 5. Common molecular classes found in petroleum.

masses or isomer sets in petroleum, lumping only the isomers at the same carbon number in the same molecular class. Additional descriptors in the SOL method could, of course, be incorporated to locate and size alkyl substitutions and branching on non-cyclic molecules should more detailed analytical information become available. If spatial orientation is crucial to the modeling of chemical processes, as in biochemical systems, more sophisticated methods of molecular descriptions could be devised.
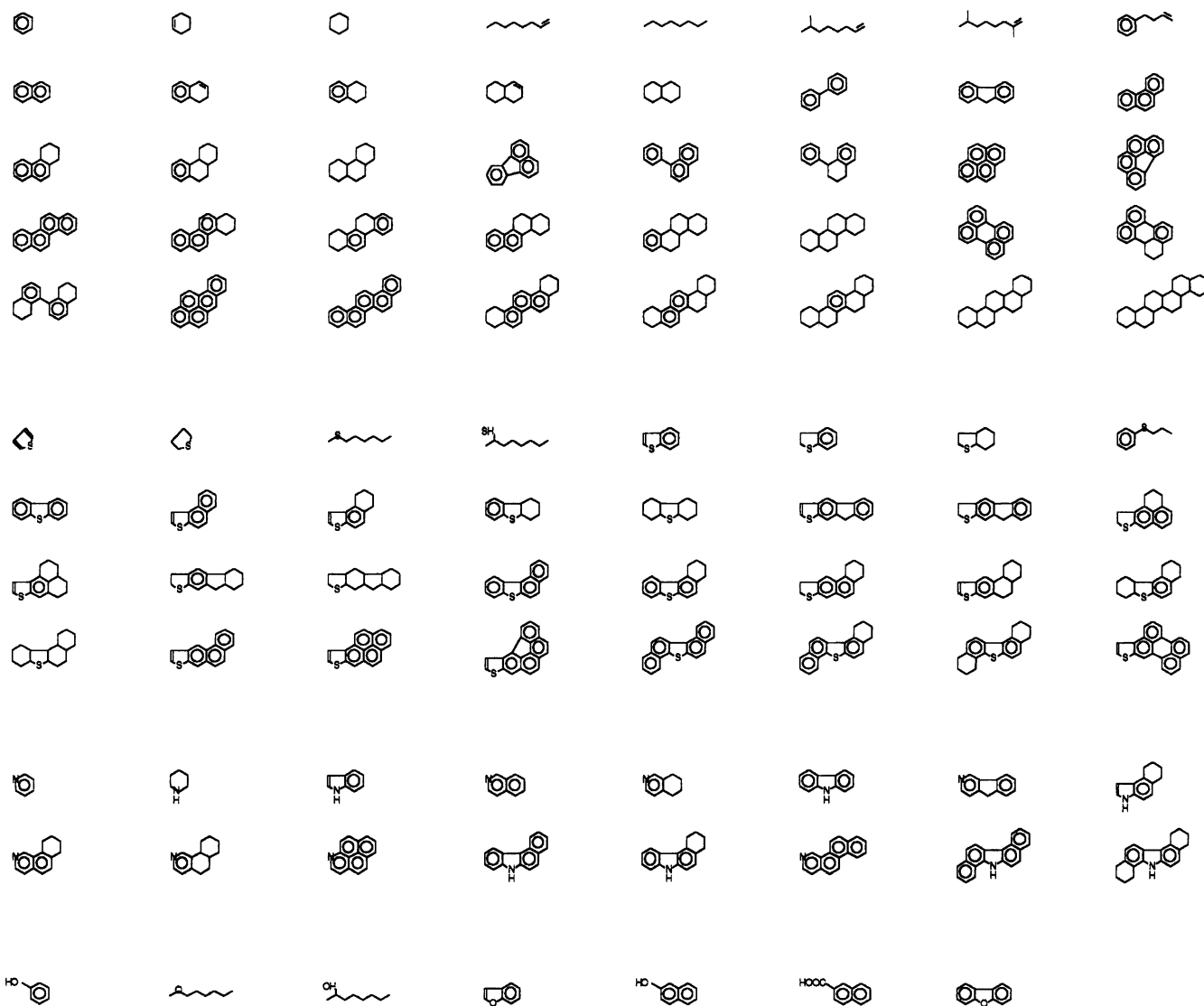
Describing the Chemistry of Refinery Processes

Why describe molecules mathematically with structural increments and why these particular increments? Although the SOL method provides a basis for organizing the molecular components of petroleum, its main purpose is to provide a convenient framework for developing a mathematical, or algorithmic, description of complex process chemistry. A characteristic feature of the catalytic and thermal chemistry of refining processes is that certain specific molecular rearrangements occur repeatedly on many different types of molecules. These particular increments correspond to those structural entities that are rearranged during reactions.

Using a limited set of structural groups to describe thousands of components enables the use of a limited set of reaction rules to establish the complex reaction networks involving tens of thousands of reactions. Each reaction rule consists of a reactant selection rule and a product generation rule. The reactant selection rule first identifies which molecules in the system can undergo a certain structural rearrangement that characterizes a particular type of chemical reaction. Logical constructs are applied to the vectors to determine which components have the increment(s) required for the reaction. The product generation rule is used to convert each reactant's vector to a corresponding product vector. Examples of reaction rules for aromatic ring saturation, ring opening, and dealkylation are given in Figure 6. For example, the aromatic saturation rule shown in Figure 6 determines if a molecule has the A4 ring required for the reaction, then converts this ring to an N4 naphthenic ring through a simple mathematical operation. The reactant vector's A4 element is decreased by one and the N4 element is increased by one to create the product vector. Molecules without an A4
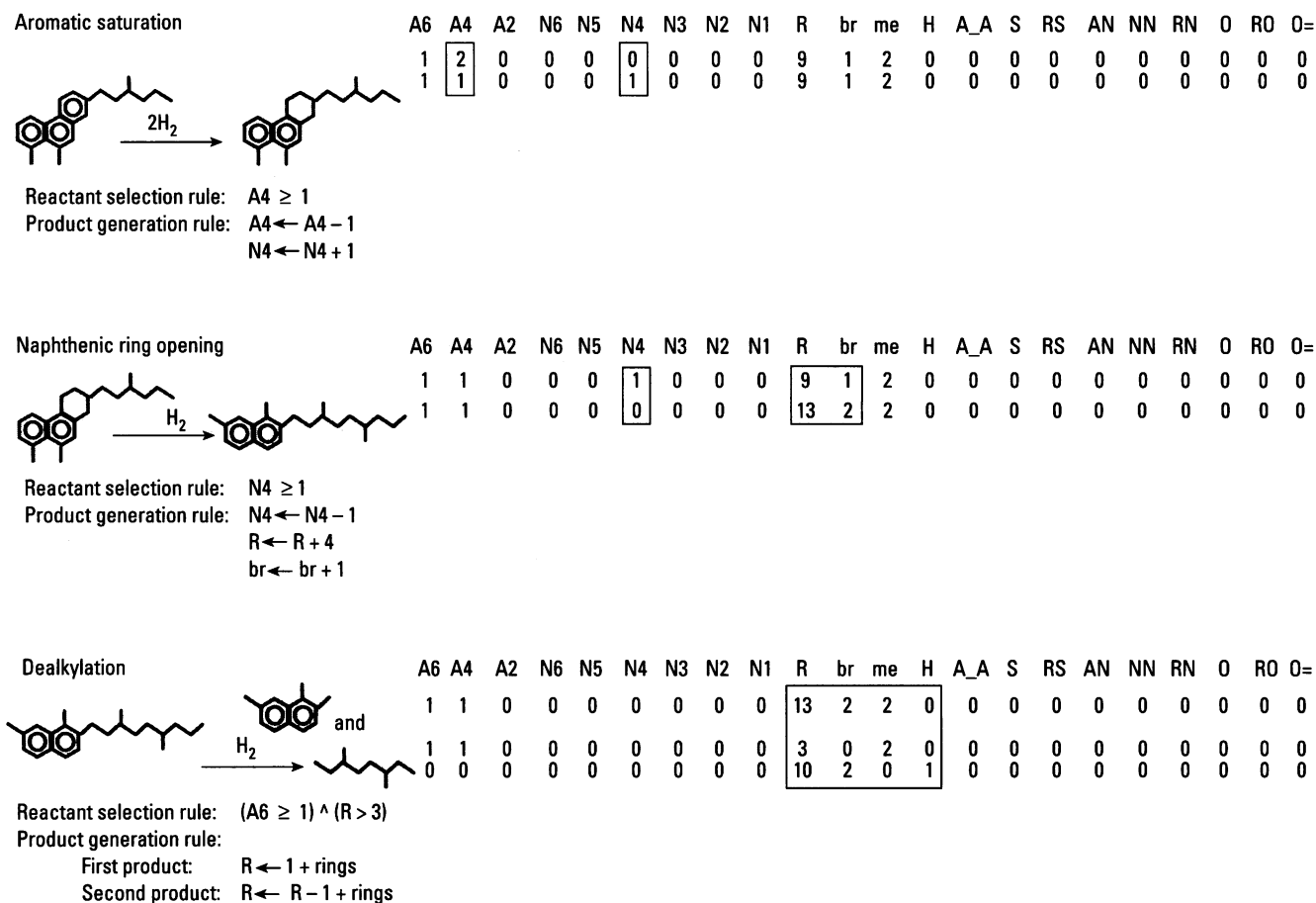
Aromatic saturation

A6	A4	A2	N6	N5	N4	N3	N2	N1	R	br	me	H	A_A	S	RS	AN	NN	RN	O	RO	O=
1	2	0	0	0	0	0	0	0	9	1	2	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	1	0	0	0	9	1	2	0	0	0	0	0	0	0	0	0	0

Reactant selection rule: A4 ≥ 1
Product generation rule: A4 ← A4 − 1
N4 ← N4 + 1

Naphthenic ring opening

A6	A4	A2	N6	N5	N4	N3	N2	N1	R	br	me	H	A_A	S	RS	AN	NN	RN	O	RO	O=
1	1	0	0	0	1	0	0	0	9	1	2	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	13	2	2	0	0	0	0	0	0	0	0	0	0

Reactant selection rule: N4 ≥ 1
Product generation rule: N4 ← N4 − 1
R ← R + 4
br ← br + 1

Dealkylation

A6	A4	A2	N6	N5	N4	N3	N2	N1	R	br	me	H	A_A	S	RS	AN	NN	RN	O	RO	O=
1	1	0	0	0	0	0	0	0	13	2	2	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	3	0	2	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	10	2	0	1	0	0	0	0	0	0	0	0	0

Reactant selection rule: (A6 ≥ 1) ∧ (R > 3)
Product generation rule:
First product: R ← 1 + rings
Second product: R ← R − 1 + rings

Figure 6. Formulating chemical reaction rules using the structure-oriented lumping representation for molecular structure. Modified from Quann and Jaffe (3).

ring are not selected for this reaction. Although hydrogen is also a reactant in this and many other reactions, it is not necessary to explicitly include it in the formulation of the rule. Information on the hydrogen stoichiometry is automatically obtained from the difference in the H content of reactant and product molecules and readily computed from the H content of the increments.

Some reaction rules, particularly those involving transformations of alkyl groups such as dealkylation or cracking, are obviously dependent on what is assumed for structure of the alkyl groups. Only experimentation with petroleum mixtures can reveal how these rules should be formulated, as it must apply to isomers having a variety of substituent structures.

Applying a reaction rule to all molecules (or vectors) in the mixture generates a reaction class for the specific chemical transformation represented by the rule. A reaction class may have thousands of reactants and their corresponding products. Each reactant has the necessary increment

for the rule. Figure 7 illustrates how reaction rules generate numerous reactant–product pairs (only a limited number are shown) for a mixture for two reaction classes.

Additional complexity arises from the fact that a molecule may satisfy the reactant selection criteria of more than one rule, i.e. a reactant may have parallel reaction pathways leading to different products. For example, the partially saturated multiring aromatic molecule in Figure 6 may undergo further saturation, ring opening, dealkylation, or even cracking of the alkyl group. Application of all rules to all vectors generates the entire complex reaction network. Application of several rules in Figure 8 to just one component and several of its offspring illustrates the generation of only a small piece of the complex network. Catalytic refining processes have 20 to 40 reaction classes or rules, resulting in over 50,000 distinct chemical transformations for the model of the process chemistry. Computer programs using sorting procedures automatically use the network to construct a reactor

model's differential rate equations for each component.

Modeling the Kinetics of Complex Reaction Systems

Constructing complex reaction networks for thousands of components and their tens of thousands of reactions creates another challenge. Each reaction requires kinetic parameters, including rate constants, activation energies and adsorption constants on catalytic sites, and chemical thermodynamic properties. Obviously, it is not feasible to conduct experimental studies of 50,000 isolated reactions. The SOL concept of reaction classes and the concept of structure–activity relationships provide a basis for determining the large number of kinetic parameters.

The first approximation for estimating reaction rate parameters is to assume that all reactions in a class have the same kinetic parameters because they undergo the same intramolecular transformation. The number of parameters required for the entire network of reactions is reduced from 50,000
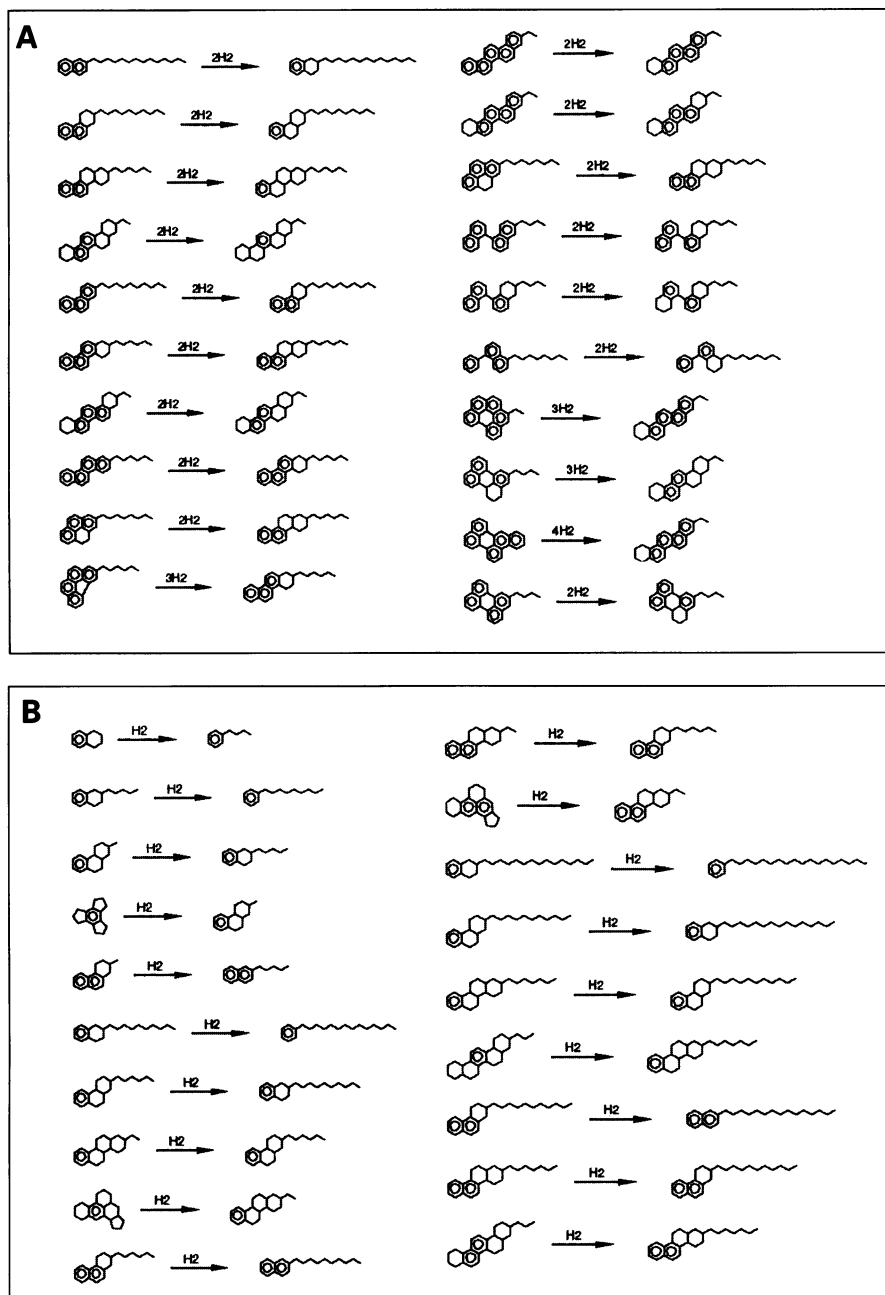
Figure 7. Application of reaction rules to complex mixtures containing many components. (A) A4 ring saturation: reactant selection rule A4 > 0; product generation rule A4 = A4 − 1, N4 = N4 + 1. (B) N4 ring cracking: reactant selection rule N4 > 0; product generation rule N4 = N4 − 1, R = R + 4.

where k_i is the rate constant for molecule i, a and b are constants to be determined experimentally, and RI_i is defined as the reactivity index for molecule i. This equation can be derived from transition-state theory and is also known as a linear free-energy relationship (LFER). The RI_i is the calculable molecular property that correlates with the free energies of the transition states of the reactants. The correct choice of RI reflects the controlling mechanism of the reaction family, such as carbenium ion stability in cracking reactions, free radical stability in thermal reactions, or electronic properties of aromatic rings in hydrogenation reactions (7). Structure–adsorption relationships have also been developed for adsorption and poisoning on catalyst sites (8). In principle, the a and b parameters and the appropriate RI can be determined from a limited amount of experimental work. The LFER can then be used to calculate the rate constant for any reaction in a large class of reactions.

The use of structure–activity relationships presumes knowledge of exact molecular structures. As discussed previously in detail, this is not the case for most of the components in petroleum or for the SOL representation. The structure–activity relationship must be developed for the isomer sets that correspond to the individual SOL components. This information can only be obtained from detailed analytical examination of products from kinetic studies on petroleum mixtures.

Calculating the Properties of Mixtures

Refinery products must meet certain quality specifications for combustion properties, fluid characteristics, or polluting potential. They include boiling point distribution, octane number, cetane number, pour point, aniline point, smoke point, cloud point, pour point, refractive index, viscosity, viscosity index, vapor pressure, surface tension, and sulfur and nitrogen content. In addition, thermodynamic properties are required for process engineering calculations and for determining the extent of reversible chemical reactions. Each mixture property is determined by the collective properties of the mixture's molecular components. Unfortunately, individual molecular properties exist only for a limited number of compounds and there is little information on the collective properties of the closely related isomers defined as an SOL component. A mixture property can be calculated if a simulated

to 20 to 40 unique rate constants and activation energies. This generally proves insufficient—within each reaction class the molecular structure of the reactant or product can influence the rate of reaction. Therefore, structure–activity relationships are required for each reaction class.

The structure–activity relationship in kinetics is not a new concept. It began with Hammet (6), who organized families of reactions to study the influence of

substituents on the rates of homogeneous reactions. The basic premise of the structure–activity relationship is that the rate constants for a family of molecules undergoing the same reaction can be correlated to a measurable or calculable molecular property. One form of this relationship is:
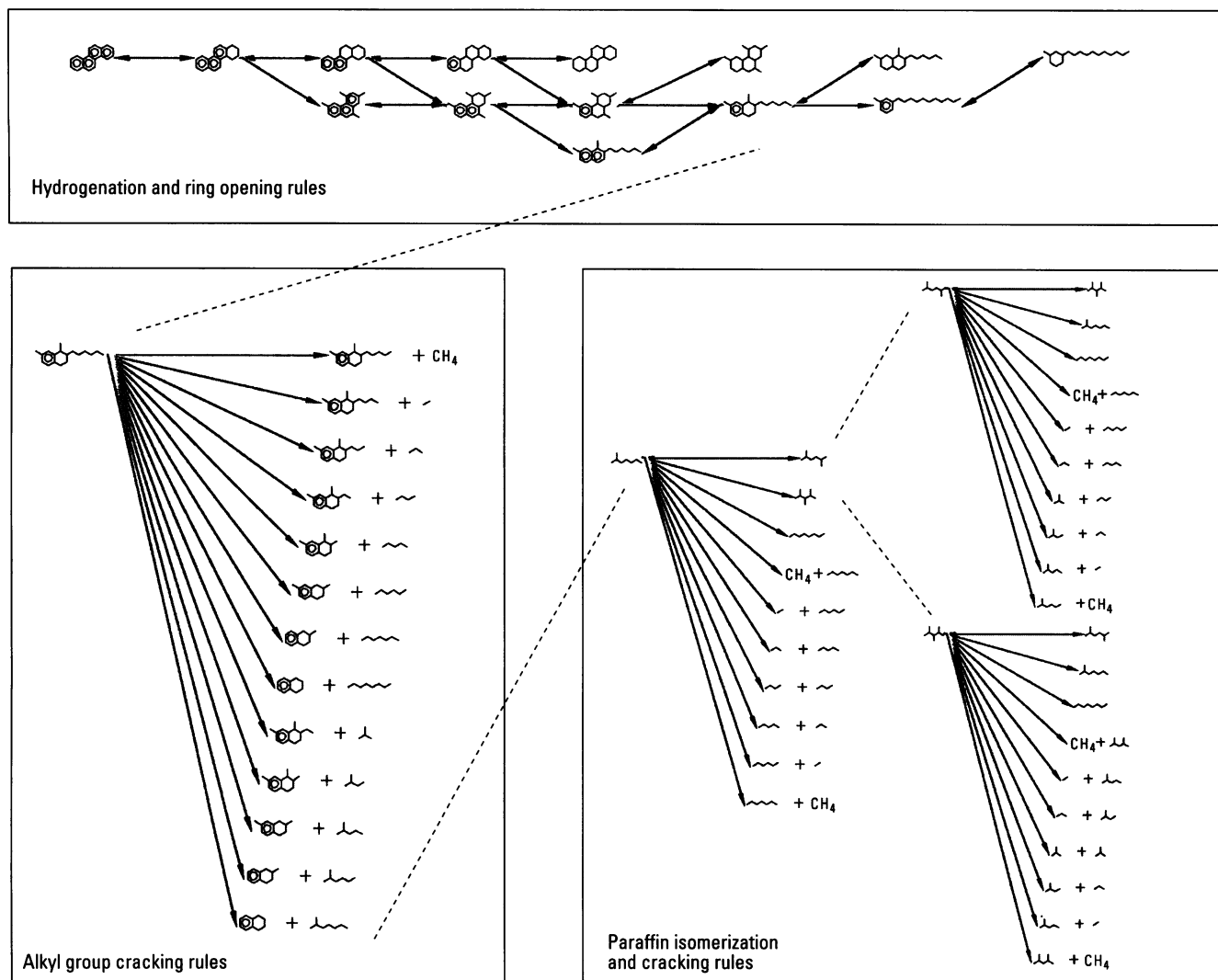
$$\ln k_i = a + b \, (RI_i)$$

Figure 8. Developing complex reaction networks through application of reaction rules.

or measured composition is available and if a structure–property correlation exists to estimate the contributions of the individual molecules or isomer sets in the mixture. There are different approaches to obtaining structure–property correlations.

One approach is to develop a new correlation for a molecular property using the molecular class and homologous series concepts as shown in Figure 9 for boiling point, specific gravity, and viscosity. The correlation relies on the property's smooth variation with carbon number for the homologous series of each molecular class. The only requirement is the availability of an extensive database on individual molecular properties. Fisher (9) surveyed various algebraic forms to represent physical properties of n-paraffins as a function of carbon number. Correlations for other molecular classes and their homologous series can be

developed as a deviation from the n-paraffin series. With the exception of the n-paraffin series, the homologous series property correlation in SOL must reflect the isomer group nature of each SOL component.

Another approach is to develop or utilize existing group contribution techniques, particularly for thermodynamic properties. Examples include Benson groups (10) for calculating molecular heat capacity, enthalpy and entropy, and universal functional activity coefficient groups (11) for liquid phase activity coefficients. Properties of molecules are computed as a sum of the properties of the groups that compose the molecules. These techniques generally have group definitions that differ from SOL and must be mapped to the SOL increment representation.

A third method for obtaining molecular properties is to use existing empirical

relationships between molecular properties. For example, critical properties for equations of state, including critical temperature, critical pressure, and the acentric factor, can be estimated from the boiling point, molecular weight, and specific gravity of the molecular components. Many of the refinery-product-quality properties can also be obtained from industry-standard correlations that use these more common physical properties.

Summary

Models of refinery processes must have the capability of predicting the yields, composition, and properties of various products. Lumping schemes with too simplistic a representation of petroleum's composition and refining chemistry have insufficient scope and reliability. A molecular-based approach to modeling is required to
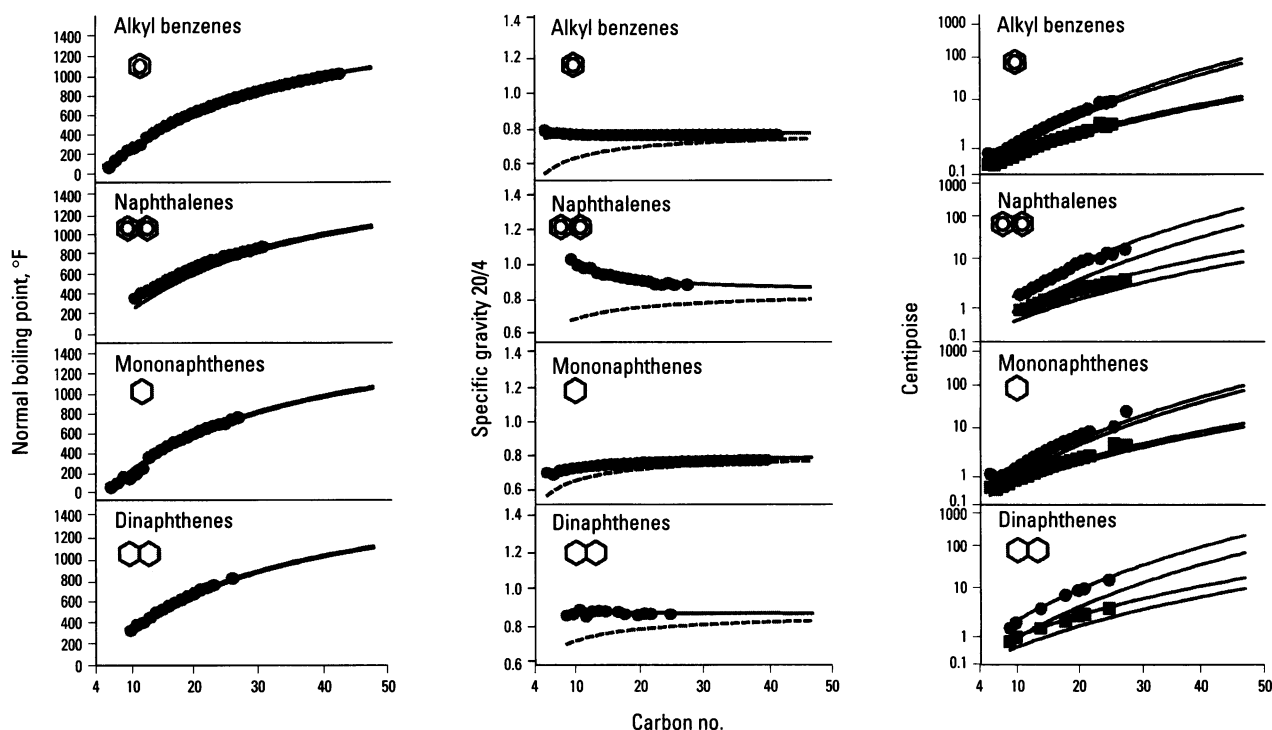
Figure 9. Correlating molecular properties using molecular class and homologous series concepts. Modified from Quann and Jaffe (3). (A) Boiling point, 760 mm, of homologous series. (B) Specific gravity of homologous series. (C) Absolute viscosity at 100 and 200°F.

capture the necessary details of mixture composition, chemistry, and properties. A molecular-based model of complex mixtures can be developed if the following criteria are satisfied. First, analytical techniques must provide some level of detail on the molecular composition and molecular structure of the mixture. Second, a method must be devised to represent and manipulate the molecular structure and chemistry of a large number of components. Third, sufficient data on molecular properties must be available to develop computational methods to estimate molecular properties. Finally, structure–activity relationships are required to estimate the kinetic parameters for a large number of reactions.

REFERENCES AND NOTES

1. Jacob SM, Gross B, Voltz SE, Weekman VW Jr. A lumping and reaction scheme for catalytic cracking. AICHE J 22(4):701–713 (1976).
2. Wei J, Kuo JCW. A lumping analysis in monomolecular reaction systems. IEC Fundam 8:114 (1969).
3. Quann RJ, Jaffe SB. Building useful models of complex reaction systems in petroleum refining. Chem Eng Sci 51:1615–1635 (1996).
4. Altgeltt KH, Boduszynski MM. Composition and Analysis of Heavy Petroleum Fractions. New York:Marcel Dekker, 1994.
5. Tissot BP, Welte DH. Petroleum Formation and Occurrence. New York:Springer-Verlag, 1994.

6. Hammet LP. The effect of structure upon the reactions of organic compounds. J Am Chem Soc 59:96–103 (1937).
7. Neurock M, Klein MT. Linear free energy relationships in kinetic analysis. Polycyclic Aromat Comp 3:231–246 (1993).
8. Ho TC, Katritzky AR, Cato SJ. Effect of nitrogen compounds on cracking catalyst. IEC Res 31:1589–1597 (1992).
9. Fisher CH. Equations correlate n-alkane physical properties with chain length. Chem Eng September 20 (1982).
10. Benson SW. Thermochemical Kinetics. New York:John Wiley & Sons, 1976.
11. Fredenslund A, Jones RL, Prausnitz JM. Group-contribution estimation of activity coefficients in non-ideal liquid mixtures. AICHE J 21:1086–1098 (1975).